Developing a Digital Human-Computer Interaction Laboratory

Lieutenant Colonel Terence S. Andre, USAF

IITA Research Publication 8
Education Series

July 2005

ABOUT THE AUTHOR

Lt Col Terence Andre is the Deputy Department Head for Research in the Department of Behavioral Sciences and Leadership at the United States Air Force Academy. He directs the department's $1.2M laboratory infrastructure and research program for faculty and cadets. He is an assistant professor in the department, teaching courses in human factors, system design, and human-computer interaction. He also oversees the human-computer interaction laboratory and directs cooperative agreements with outside research agencies. Lt Col Andre received his commission in 1987 after graduating from the United States Air Force Academy. He entered the Air Force in 1987 at Williams AFB, where he began his initial work as a human factors scientist. He had tours at Vandenberg AFB and Kirtland AFB as a program manager and human factors evaluator. He was recently the branch chief for the Warfighting Training Research Division at the Air Force Research Laboratory in Mesa, AZ where he led the division's $10M research program in Distributed Mission Training. Lt Col Andre received his Masters in Industrial Engineering from Cal Poly and his PhD in Industrial and Systems Engineering from Virginia Tech.

Comments pertaining to this report are invited and should be directed to:

Sharon Richardson
Director of Conferences and Publications
Institute for Information Technology Applications
HQ USAFA/DFEI
2354Fairchild Drive, Suite 4L35F
USAF Academy CO 80840-6258
Tel. (719) 333-2746; Fax (719) 333-2946Email:
mailto:sharon.richardson@usafa.af.mil

Table of Contents

List of Tables

List of Figures

Developing a Digital Human-Computer Interaction Laboratory

ABSTRACT

The Behavioral Sciences and Leadership Department at the United States Air Force Academy (USAFA) recently initiated an effort to develop a low-cost usability evaluation system for undergraduate education and research. Based on student input, we knew we needed a flexible and portable system that would be cost effective for both data capture and analysis. Our overarching goal was to develop a system that would be easy for students and faculty to learn and maintain. In addition to creating a system that would be flexible, portable, and easy to learn, we wanted to develop a learning environment around that system for undergraduate students in human factors, computer science, and systems engineering. This paper documents the process we followed to design and implement our lab, and provides a step-by-step solution for developing similar low-cost usability laboratories at other universities, both for teaching and research. By integrating a software-based usability recording tool (Morae™) as the main component of the laboratory, we were able to develop the solution we needed and provide cadets at the Air Force Academy with the same capability as high-end laboratories. We plan to integrate other methods and tools in the future to support efficient usability diagnosis and evaluation for university faculty and students.

INTRODUCTION

Usability evaluation has become quite popular in industry over the past twenty years, with organizations spending huge sums of money to build state-of-the-art laboratories (Hix & Hartson, 1993). In the last few years, focus has shifted to low-cost usability evaluation facilities with the desire to have the same high-end capability of more expensive labs. The motivation to conduct usability evaluation has remained strong and no longer requires justification in most organizations. The nature of today's web-based and desktop software requires usability evaluation to remain competitive (Butler, 1996). Similar to corporations, universities have invested in the process of developing usability laboratories. This is especially true in computer science, industrial engineering, and psychology graduate programs at major universities, where the focus is on conducting research in usability evaluation methods.

Usability laboratories at these universities have often mirrored the look and feel of laboratories in industry, albeit on a much smaller scale. Graduate programs with usability laboratories have contributed to the basic methodology and theory of usability evaluation, which has been adapted and refined by industry. Because the technology used in graduate programs has typically reflected the current state-of-the-art in audio and video recording, usability laboratories at these programs have tended to be very sophisticated, requiring extensive hardware in the form of video cassette recorders (VCRs), audio and video mixers, camcorders, scan converters, and extensive cabling. In addition, many university usability laboratories have built rooms with one-way mirrors so one or more observers could easily monitor testing sessions.

The downside to the hardware-intensive layout of usability laboratories at universities has been the expertise required to maintain the equipment from year to year. As technology has improved, many programs have found it cost-prohibitive to upgrade equipment in a university laboratory. In addition, a generally high turnover of personnel who are experienced with the equipment has often resulted in a "continuity gap" for effectively operating the equipment. As a result, many usability laboratories at universities have ended up sitting idle because the equipment had to be frequently modified to the point where no one knew how to restore it to the original configuration.

Usability laboratories in universities have typically not been designed as "walk-up-and-use" systems because of the advanced audio, video, and screen capture functions that needed to be synchronized. As a result, hundreds of hours of VCR tapes often end up in storage with the intention of further editing and analysis. However, the analysis never happens because it is common to spend 3-10 hours editing each hour of video-based usability data (Nielsen, 1993), making it unfeasible.

The above dilemma motivated the Behavioral Sciences and Leadership Department at the USAFA to develop a very low-cost, all-digital software usability evaluation laboratory for education and research. We faced the challenge of developing a flexible and portable system that would be cost effective for both data capture and analysis. Our primary goal was to develop a system that would be easy for students and faculty to learn to use and maintain. In addition to offering flexibility, portability, and ease of use, we also wanted this system to foster a learning environment for undergraduate students in human factors, computer science, and systems engineering.

GOALS AND OBJECTIVES

The overall goal of the Digital Human-Computer Interaction Laboratory (HCIL) project was to develop a low-cost usability evaluation system for undergraduate education and research. To accomplish this goal, the HCIL Project Team defined several objectives which helped shape the philosophy of our design and research. These objectives were:

- Integrate use of commercial-off-the-shelf hardware and software wherever possible
- Use existing laboratory facilities with no physical modifications to rooms
- Build in flexibility so the lab can be modified easily at minimum cost as hardware and software change
- Create a system that is easy to learn how to use and easy to maintain by cadets and faculty

- Provide a digital storage environment to eliminate the need for analog recording equipment like VCRs
- Create a teaching laboratory where contemporary usability evaluation methods are easily integrated so that cadets can learn about the field while they are using the methods

All of the above objectives have been accomplished as of the writing of this report. The last objective will continue to foster further research in the HCIL as new methods are integrated and tested.

PROJECT TEAM ORGANIZATION

The Digital HCIL Project Team involved several different mission elements at USAFA and collaborators from industry. The list below identifies the key organizations and individuals on the HCIL Project Team.

- ☐ IITA
  - General James P. McCarthy, (Retired), Director
  - Lt Col Jim Harper, Managing Director
  - Lt Col Ellen Fiebig
  - Dr. Eric Hamilton
- ☐ DFBL
  - Lt Col Terence Andre, Project Manager
  - 2Lt Austen Lefebvre, Lab Resource Manager
  - Mr. Randy Torres, Lab Director

- ☐ DFET
  - Mr. Rob Wells, Director, Academic Media

- ☐ 10 CS
  - Mr. Mark Wellauer, Superintendent
  - Mr. Dilver Brown
  - Mr. Edward Voltz
  - Mr. Alexander Zehnder

- ☐ Cadet Wing
  - C1C Paul Doran
  - C1C Julie Baker
  - C1C Ryan Herman
  - C1C Christina Williams
  - C1C Apphia Taylor

- ☐ Air Force Research Laboratory (Mesa, AZ)

- Dr. Winston Bennett

☐ TechSmith (Okemos, MI)
- Mr. Shane Lovellette

☐ Virginia Tech
- Dr. Rex Hartson
- Mr. Jon Howarth

DEVELOPMENT PROCESS

Our development effort began during the Spring 2004 semester with cadets at the USAFA enrolled in a human factors design course. The cadets were studying the design process and implemented a strategy originally defined by Williges, Williges, and Elkerton (1987). Williges et al. noted the importance of using a systematic process for conducting Human-Computer Interaction (HCI) research. Our process at the USAFA mirrored the design stages outlined by Williges and Hartson (1986) and Williges et al. (1987) and included the following three stages: (1) initial design, focusing on requirements and specifications; (2) formative evaluation to examine if early concepts are moving closer to accomplishing the goals; and (3) summative evaluation using experimental procedures to compare to other methods. These stages provided the process for accomplishing the objectives of the project.

Cadets in the human factors course researched various industry and academic HCI labs on the internet, and then visited several labs in Virginia and Maryland to gain a perspective on the necessary requirements. Cadets were able to see HCI facilities at Virginia Tech in Blacksburg, Virginia, the Census Bureau in Washington DC, and UserWorks in Silver Spring, Maryland. These facilities gave the cadets a good perspective of academic, government, and private industry capabilities for HCI laboratories. Based on these site visits and further consultations, we developed the goal of creating a PC-based usability analysis environment with the following objectives:

- Ability to observe, record, and annotate from local or remote machine with minimal distraction to the research participant
- Embedded analysis capability to examine user performance (task, errors, keystrokes, mouse clicks, web page changes, and such.)
- Digital storage of recording sessions (i.e., eliminating the requirement for VCR editing)
- Maximum use of commercial-off-the-shelf hardware and software
- Maximum use of existing laboratory facilities with no physical modifications
- Supports flexibility in implementation that can be modified easily at low cost as hardware and software change at a low cost

Our development of the usability evaluation laboratory also considered the importance of integrating the capability as a teaching laboratory within the human factors and systems engineering programs at the USAFA. We wanted to create an environment where the main observation room provided unobtrusive monitoring of research rooms while a class section observed usability evaluation techniques in real time.

We established two phases for the development of the usability evaluation laboratory. During the first phase, cadets and faculty focused on establishing a local recording environment between existing adjacent laboratory rooms. The goal was to integrate desktop computers, local area network connections, internet conferencing software, and inexpensive web cameras to establish a very simple recording environment. During the second phase, our goal was to expand the scope to include an observation room where a “teaching laboratory” could be established. Our focus in the observation room was to provide the evaluator with the capability to observe usability evaluation sessions from one of two different participant rooms.

INTEGRATION OF SOFTWARE-BASED RECORDING TOOL

During the development process, we established a Cooperative Research and Development Agreement (CRADA) with the TechSmith Corporation based in Okemos, Michigan. Through the CRADA we were able to use a beta version of a new usability evaluation tool called Morae™. Morae consists of three components that record and synchronize user actions with detailed application and computer system data for the analysis of human-computer interaction (Morae Overview Whitepaper, 2004). It provided us an all-digital solution to record usability sessions without the use of traditional hardware recording and editing equipment. Modeled on commonly accepted usability testing processes, Morae did not require major changes in our usability testing methodologies, and its design offered several advantages that helped accomplish our objectives (see Table 1).

Morae consists of three components that can be configured in different ways to conduct testing and analysis: Morae Recorder, Morae Remote Viewer, and Morae Manager. By separating the recording, observation and logging, and analysis and presentation processes into separate components, Morae provided us the flexibility we needed to set-up a lab within our existing facilities. Additionally, the multiple component structure enabled us to create a portable usability testing lab that we use for field research. It consists of a laptop with Morae Recorder installed and a Web camera.

USAFA OBJECTIVE	MORAE DESIGN
Observe, record, and annotate from local or remote machine with minimal distraction to research participant	Single solution for recording, observing and logging, analyzing, and sharing usability tests without disturbing the participant
Examine user performance (tasks, errors, keystrokes, mouse clicks, Web page changes, etc.)	Automatic capture of participant interaction data synchronized and time-stamped, including mouse clicks, keystrokes, Web page changes, etc.
Digital storage of recording sessions	Records video, audio and data in digital format
Maximum use of commercial off-the-shelf hardware and software	Off-the-shelf software
Maximum use of existing laboratory facilities with no physical modifications	Utilizes existing network infrastructure
Flexible and low-cost system that can be easily modified as hardware and software changes	Three components support various configurations and they are upgraded regularly with new functionality compatible with hardware advances

Table 1: Integrating Morae Capability with Digital HCIL Design Objectives

Morae Recorder

Recorder is the data-collection component of Morae that runs on the computer the test participant interacts with. Because it runs silently in the background, it met our requirement of not disturbing the participant during the test session. Recorder automatically captures both media and participant interaction data during testing. Examples of what is captured are shown in Table 2.

Media Captured	Interaction Data Captured
• Video of the screen • Video of the participant through a Web or Digital Video camera • Audio of the participant	• Keystrokes • Mouse Clicks • Web page changes • Text appearing on screen • User Interface events (i.e. Window, menus & buttons getting focus or being resized)

Table 2: Information Captured from Morae Recorder

The media and interaction data are synchronized automatically, which was a major advantage for our lab, because it saved both time and resources by eliminating the need to manually synchronize participant and screen video. Achieving synchronization without Morae would have required multiple pieces of expensive hardware components. Additionally, since the interaction data is automatically time-stamped and indexed to the media, we didn't have to dedicate observers to log those events manually.

Morae Remote Viewer

The Remote Viewer is the observation and logging component of Morae. It enables one or more observers to watch a usability test live over a network (using a LAN or broadband connection) from a remote location. To do this, the Remote Viewer connects to the Recorder component and provides observers the option to view the screen video of the test participant in real-time, or to stream the screen video, camera video (as a picture-in-picture) and audio with a short buffering delay (typically 8-10 seconds). One or more Remote Viewers can connect to a Recorder component simultaneously from different locations (see Figure 1).

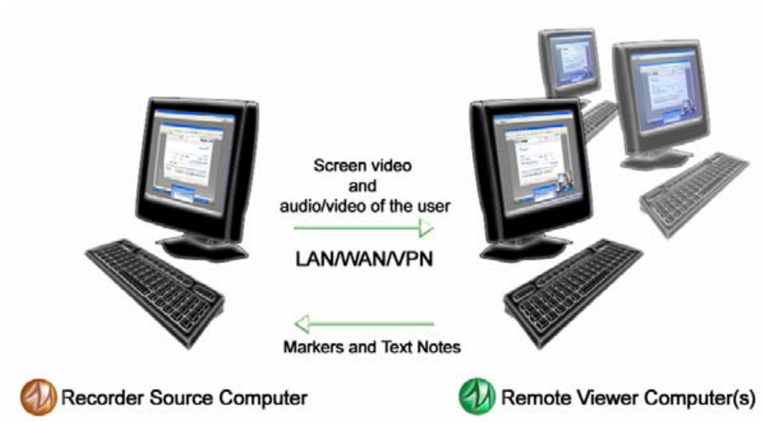

Figure 1: Recorder – Remote Viewer(s) Connection

The Remote Viewer component enabled us to take advantage of our existing LAN and building facilities with minimal modifications. We were able to share the screen video from two different participant rooms in real-time over our LAN. Since we were creating a learning environment, we needed a method for the cadets to watch usability testing sessions without being in the same room as the participant. With the Remote Viewer component, cadets and faculty can watch a testing session from any LAN connected location. This eliminated the need for one-way glass and expanded the number of observers possible.

The other advantage provided by the Remote Viewer is the ability to set markers and add associated text annotations, which are communicated to Recorder, synchronized and saved. This enabled our cadets and researchers to log a test from any location. Since the Recorder automatically captured interaction data, cadets used the markers in Remote Viewer to focus on logging qualitative, participant-based observations, such as when the test participant became frustrated, asked for help, or got confused (see Figure 2).

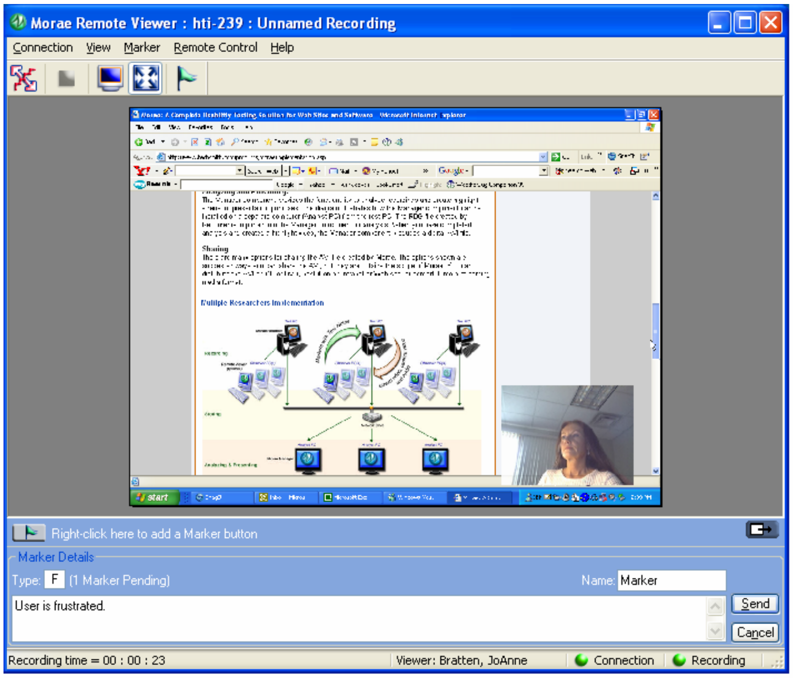

Figure 2: Morae Remote Viewer Interface

Morae Manager

The Manager is the component of Morae that provides analysis and presentation capabilities. As described earlier, one major bottleneck of usability testing has been the inordinate amount of time necessary to analyze video, typically 3-10 hours per hour of video recorded. Because of this, video analysis just hasn't been done, limiting the amount of useful analysis to only what is gleaned from live observation of usability tests. In a teaching environment like ours, cadets just don't have the time to dedicate to video analysis, yet they need to be able to go back and review test sessions to better understand where issues exist and how to recognize them.

Because Morae indexes the media (screen video, camera video and audio) with interaction data, the time required for video analysis is greatly reduced. In Manager, for example, cadets are able to search a usability test for all of the mouse clicks that occurred. Manager displays a list of all clicks with metadata information related to each one (when it occurred, what application it occurred in and which mouse button was clicked) as shown in Figure 3. By selecting one of the clicks, the screen and camera video move to the point in time when that click occurred and highlight where on the screen it occurred. Manager supports searching for any of the interaction data or observer markers captured by Recorder.

This method of searching the video to find specific events gives our lab a great advantage. Cadets don't have to spend time fast forwarding and rewinding a videotape, looking for events of interest. These events are quickly accessible and they enable cadets to further review and analyze the data by watching and listening to the participant several times, as needed. Additionally, calculating time on task, navigational path, number of clicks to complete a task, error rates, success rates, and other metrics are supported. The interaction data can also be exported to a comma-delimited format which can be opened in Excel or another statistical program for further manipulation and analysis, if needed.

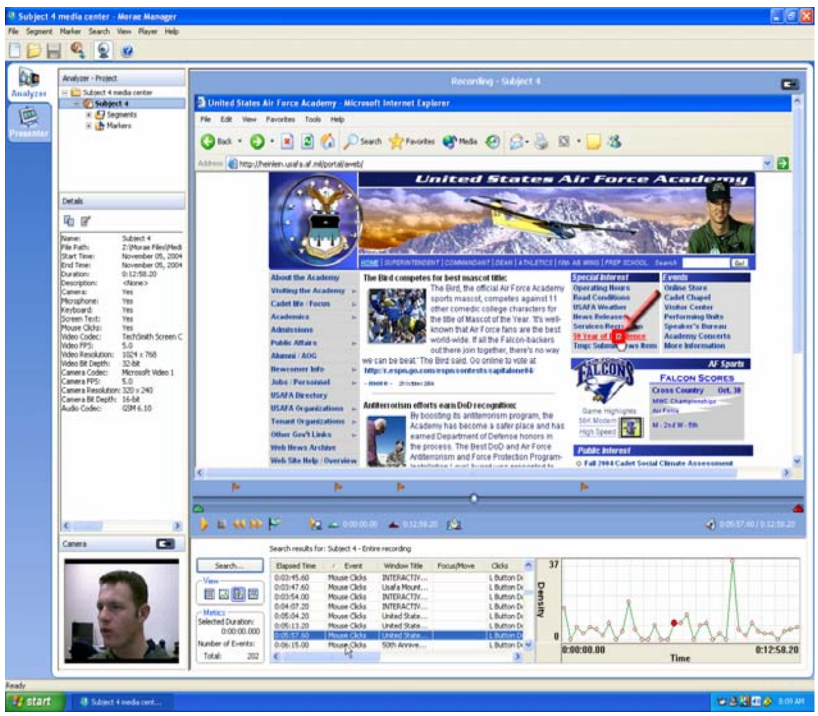

Figure 3: Morae Manager Interface with Search Result Highlighted

Another advantage provided by the Manager component is the ability to easily create a highlight video from usability tests. Highlight videos are often shared with designers, developers and managers to demonstrate where issues exist with a user interface (UI) or Web page. Seeing actual users interacting with software or Web sites is very powerful and reinforces the analysis. In academic environments, the highlight videos have the additional purpose of being a learning tool, both in their creation, when students learn hands on how to identify problems when creating the clips and when viewing them, and students can compate analysis results fro peers and faculty.

In the past, creating these videos required several different computer software and hardware configurations, which were difficult to learn and use. The Manager component's simple, integrated editing interface eliminates the need for additional hardware and software, which not only saves money but also reduces complexity.

By integrating Morae into our lab, we were able to reach our goals of creating a flexible, portable and easy-to-learn system that creates a learning environment for both students and faculty. As an off-the-shelf application, Morae made it possible for us to utilize our existing computers and facility infrastructure with minimal changes. Additionally, we were able to greatly reduce the amount of equipment necessary to operate the lab, which saved us time, money and resources.

LABORATORY SETUP

One of our objectives was to use our existing laboratory facilities with no physical modification to the rooms to create an observation environment. That is, we did not want to use one-way mirrors between rooms, because that did not provide the flexibility of moving the lab capability around if needed. We identified three rooms where we could stand up the usability recording environment. Since we focused primarily on using the local area network for most of the recording burden, we could locate our main observation room just about anywhere. The main requirement was to have a room big enough for a class of 15-20 cadets to observe a session. We also wanted our subject rooms to accommodate a single user or maybe a team of cadets working on an application such as a command and control task.

During the first phase of our work, we focused our efforts on establishing a local recording and observation area in one room. The local recording/observation room included a PC for the participant (the recording PC) and a dual-screen PC for the observer (the observation PC). The initial capability used only a single web camera connected to the recording PC. Figure 4 shows the original room we established for local recording/observation using Morae (Subject Room 1).

With the first room established, we moved into the second phase, which involved adding another recording room (see Figure 5) and the Observer Room (see Figure 6). We installed digital camcorders and small security cameras in each of the recording rooms in order to have constant video feeds into the Observer Room. The digital camcorders were used as the primary recording devices for integrating user video via Morae Recorder while the security cameras provided the context of the entire room, in case we were interested in observing team-based tasks.

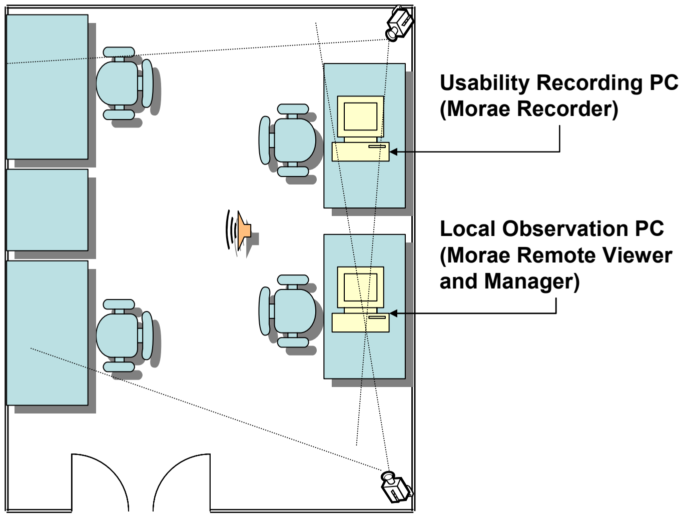

Figure 4: Subject Room 1

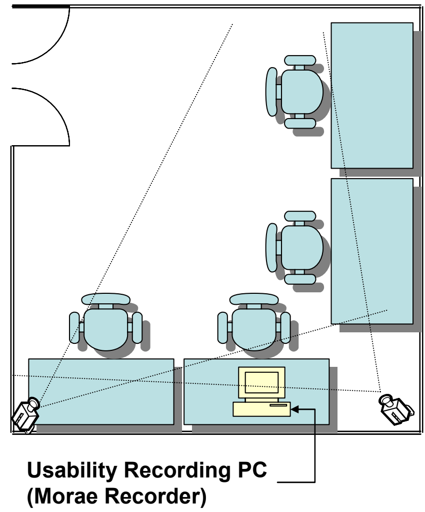

Figure 5: Subject Room 2

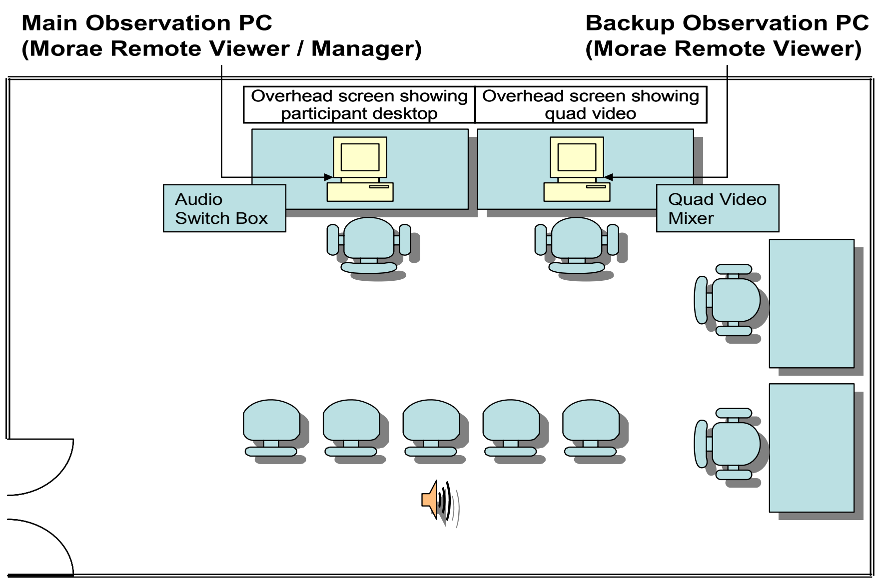

Figure 6: Observer Room

In the Observer Room, we built a "teaching laboratory" that enabled cadets to observe usability sessions from Subject Room 1 or 2. We established the primary connection between the Observer Room and Subject Room 1, which consisted of participant and observer microphones, audio speakers, and an audio switch box, in order to create an intercom system and real-time, high-fidelity audio/video recording environment. To support the teaching laboratory, we projected the screen video of the participant's desktop on one screen and the live camera video from the Subject room on an adjacent screen as shown in Figure 7. The diagram in Figure 8 shows how we built the audio connections between these two rooms to enable the observers to communicate with the participant or make direct audio comments to the Morae Recorder software without the participant hearing our comments.

Figure 7: Observer Room with the Desktop Screen Video (left) and Real-Time Camera Video (right)

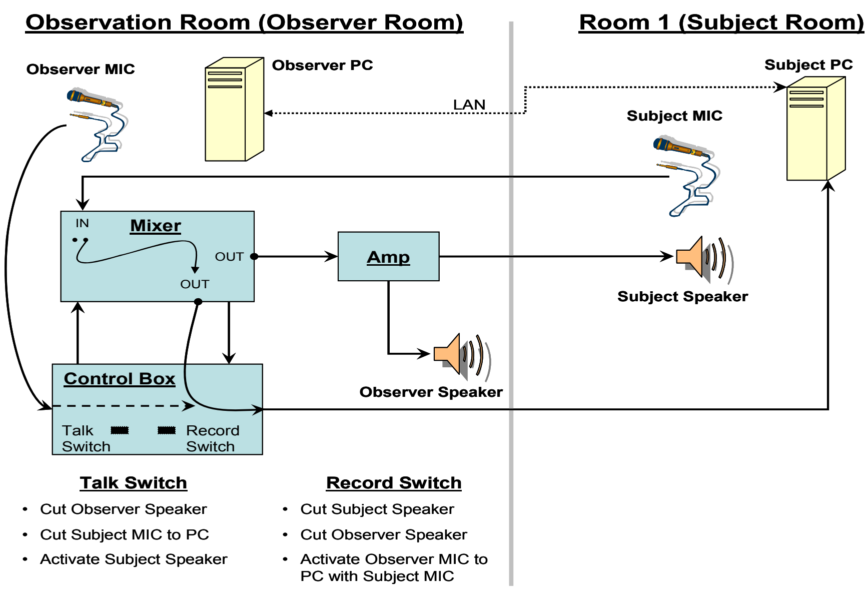

Figure 8: Audio Connections between the Observer Room and Subject Room 1

Another objective was to have a flexible system that would take advantage of existing low-cost hardware and software solutions so that a basic, portable solution could exist anywhere in our laboratory. We decided to incorporate a backup observation workstation that relied solely on Morae's software-based solution without the high-fidelity audio or video connections we implemented with the main observation workstation. Our goal was to develop "audio/video anywhere" through the use of Morae and Microsoft NetMeeting software. NetMeeting provided a way to set up an audio/video conference between two PCs that are connected on a LAN. With NetMeeting, we were able to develop an intercom system that provided a way to receive live audio and video from the participant as well as send live audio to the participant room (for directing the participant and providing task completion feedback). Combined with Morae Recorder on the participant PC and Morae Remote Viewer on the observation PC, we were able to conduct a real-time usability evaluation session from our backup observation workstation while another session was being conducted with the main observation workstation. When we didn't need a real-time audio/video session, we could use either workstation to stream the screen video, camera video, and audio with a short delay using Morae's streaming option in the Remote Viewer. Morae's streaming capability provided us with the most flexibility and portability with the smallest "footprint" of required software/hardware.

EQUIPMENT COSTS

When we set out to establish a usability evaluation laboratory at the USAFA, we wanted to create an HCI teaching laboratory with a software-based solution that was low-cost and flexible enough to meet our changing needs. We ended up creating a dual-use environment; a high-fidelity solution for teaching cadets HCI tools and methods, and a low-cost solution that can be run from anywhere in our laboratory, or even anywhere in the USAFA. Table 3 summarizes the equipment we acquired for our laboratory and shows the comparison between the high-fidelity teaching laboratory and the low-cost flexible laboratory, which requires a minimal set of hardware when using Morae and NetMeeting for conferencing. As shown in Table 3, we were able to develop a high-fidelity solution for approximately $7,000. The low-cost solution using NetMeeting was accomplished for approximately $1,400. These figures do not include the cost of PCs, because we were able to use existing systems in our laboratory. Most universities have existing PCs that are adequate for running the software and storing the data files. Data storage is probably one of the most important upgrades for anyone using a digital recording system and is relatively inexpensive today, with hard drive prices averaging $1 per gigabyte.

Equipment Items	Location	High-Fidelity Teaching Laboratory with Wired Audio/Video	Low-Cost Laboratory using NetMeeting for Intercom
Morae License		≈ $1100 (Academic)	≈ $1100 (Academic)
Digital Video Camcorder	Participant	≈ $700	
Color Security Cameras (2)	Participant	≈ $420	
Cardioid Microphone	Observer	≈ $170	
Boundary Microphone	Participant	≈ $190	
LCD Projectors (2)	Observer	≈ $3200	
Audio Switch Box	Observer	≈ $190	
Mixer and Audio Rack	Observer	≈ $500	
Amplifier	Observer	≈ $175	
Room Speakers	Both rooms	≈ $100	
Cables and wall plates	Both rooms	≈ $200	
USB Web Camera (Morae Recorder)	Participant		≈ $100
USB Tabletop Microphone (NetMeeting)	Participant		≈ $40
USB Headset Microphone (NetMeeting)	Observer		≈ $75
TOTAL		**≈ $6945**	**≈ $1315**

Table 3: Equipment Inventory for High-Fidelity and Low-Cost Laboratory Solutions

RESEARCH STUDY

Cadets have been involved in the Digital HCIL Project since its inception. In addition to their help in the conceptual design, cadets have used the lab in classroom and independent research projects. One of the research questions explored by a group of cadets concerned the usefulness of nonverbal cues from video captured during usability evaluation. That is, does video data of nonverbal cues help usability experts more accurately detect usability problems than data sets with audio alone? The Digital HCIL allows the evaluator to automatically mix the desktop screen, user audio, and picture-in-picture (PIP) of the user into one video file.

Research has shown in other contexts that nonverbal cues provide information that is nearly impossible to detect from a verbal protocol. According to Patterson (1983), nonverbal cues such as observed behavior are more representative of the true characteristics, feelings, and attitudes of a person. Nonverbal behavior is often unconscious and sincere, while the verbal output of an individual is more conscious and easily manipulated to sound as the user believes necessary (Patterson, 1983). Previous research has indicated that nonverbal cues can enhance verbal communication which is used through a participant's introspection of his/her performance on a designated task with the program (Argyle, 1972; Argyle & Dean, 1965; Argyle, Lalljee, & Cook, 1968; Kendon, 1967). Furthermore, one of the most basic functions of nonverbal cues is providing information that is otherwise non-existent without the use of video imagery. For these reasons, cadets were interested in the importance of nonverbal cues as they are used in usability problem identification.

This experiment used the Morae software to create two highlight films of novice users performing tasks on the Internet Movie Database (IMBD). IMBD was chosen due to its multitude of functions and its unknown reputation to the public. Nine participants were brought in individually to perform a search task with the IMDB site.

The participants were asked to think aloud while completing the tasks. The participants were told to act as if the experimenters were behind a wall and that they could not see what the participants were doing on the computer. The participants needed to tell the experimenters exactly what task they were doing and how they were going to complete the task, articulating every action they performed and every thought they had when interacting with the website.

Highlight video clips of a representative sample of participant actions were produced using the Morae software. The highlight video clips were produced in two formats. One format contained video of the screen, user audio, and picture-in-picture (PIP) of user video (PIP group) and the other included just video of the screen and user audio (no-PIP group). Figure 9 shows an example of the PIP stimulus set while Figure 10 shows an example of the no-PIP stimulus set. The two highlight video clips were then shown to 24 human factors students in a classroom setting. Twelve students analyzed the PIP version and twelve analyzed the no-PIP version.

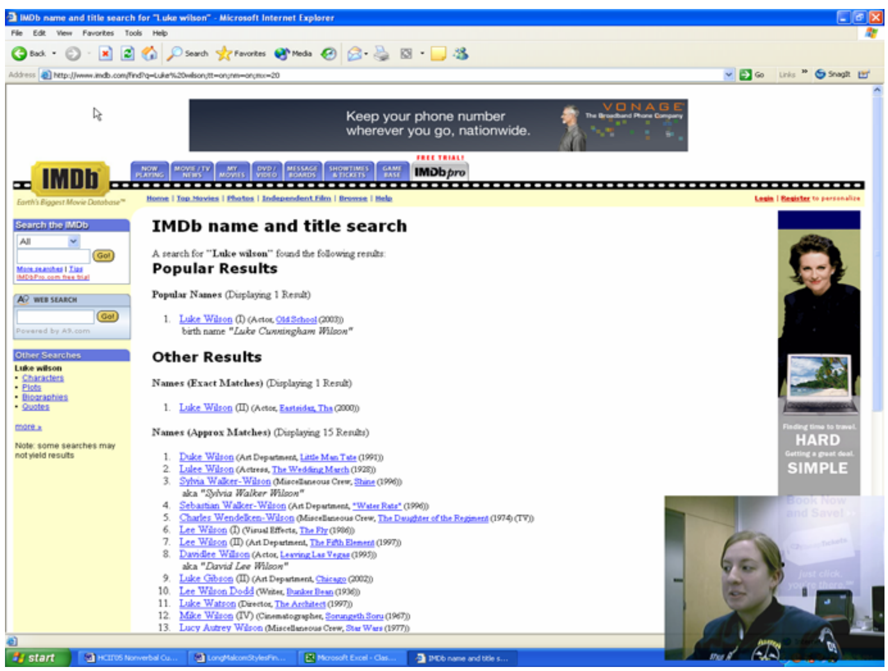

Figure 9: Example of PIP Video Clip

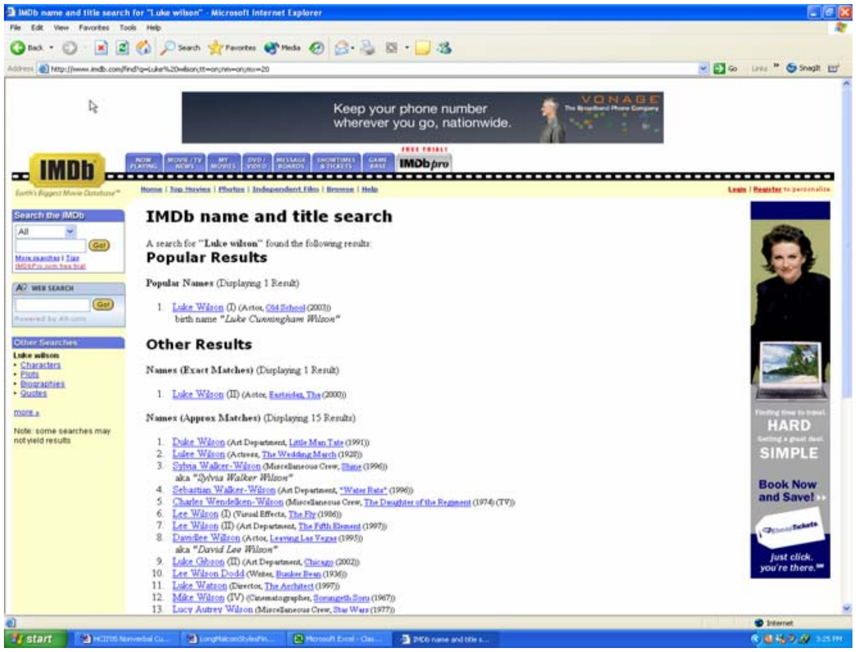

Figure 10: Example of No-PIP Video Clip

Students from the human factors class completed worksheets identifying usability problems as they watched the video clips. All 24 students received the same training on how to identify a usability problem.

Figure 11 shows the mean number of usability problems found by the students in the PIP group (M = 3.75) versus the no-PIP group (M = 3.67). This difference was tested using an independent groups t test, and was shown to be nonsignificant, $t(22) = .192$, $p = .850$. Figure 11 clearly shows that the two groups found approximately equal number of problems on average. More variability appears to exist in the no-PIP group as noted by the higher standard deviation (SD = 1.23) as compared to the PIP group (SD = 0.87). The total number of unique problems found in the PIP group was 7 compared to 11 in the no-PIP group.

We also looked at the level of agreement among student evaluators using Cohen's Kappa (Cohen, 1960). Using the Kappa statistic, the level of agreement for the PIP group was 0.25 ($p < .001$) while the agreement for the no-PIP group was 0.35 ($p < .001$). These results show that there is a "fair" amount of consistency in determining errors that would be expected by chance. The "fair" rating is based on recommended values by Landis and Koch (1977)* where the level of agreement is between .2 and .4.

*The measurement of observer agreement for categorical data:
Poor agreement = Less than 0.20

Fair agreement = 0.20 to 0.40
Moderate agreement = 0.40 to 0.60
Good agreement = 0.60 to 9.80
Very good agreement = 080 to 1.00

Landis, J.R. & G.G. Koch, *Biometrics* 33,159-174.

Figure 12 shows the results from a survey given to each group at the end of the experiment. The question asked the PIP group if they agreed that having a picture-in-picture video of the user helped them identify usability problems. A similar question was given to the no-PIP group, asking them if it would have been helpful to have picture-in-picture video of the user included in the video clip. The students answered their respective question using a 5-point Likert scale, with 1 being "Strongly Disagree" and 5 being "Strongly Agree." Students who received the PIP recordings reported moderate agreement (M = 3.75) in terms of the user video helping them identify usability problems. The no-PIP group reported slightly stronger agreement (M = 4.08) that having user video available to them would have been useful in identifying usability problems.

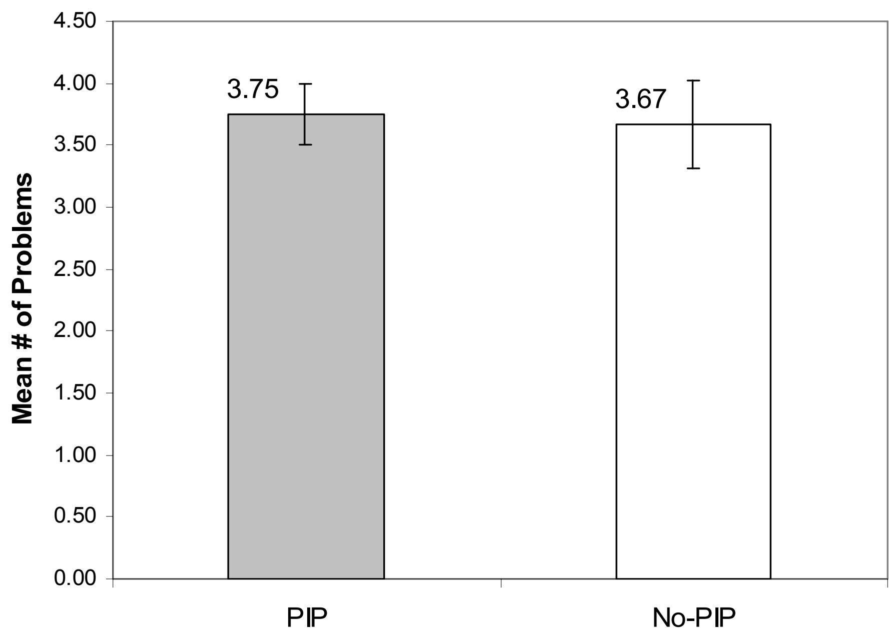

Figure 11: Mean # of Problems Found for each Group

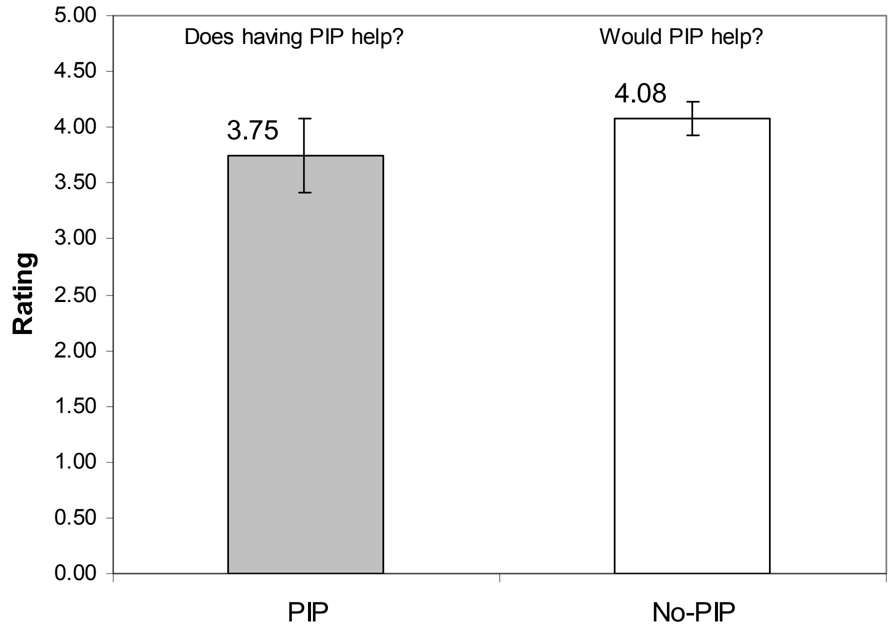

Figure 12: Rating of Whether Picture-in-Picture Video of User is Perceived to Help or Not

The cadet research study showed that including user video does not significantly increase or decrease the number of usability problems identified on average. Factors that could impact the number of problems identified for either group might include how much the user "talks" about their interaction experience and the experience of the evaluator. This study did not show agreement level to be a conclusive metric for determining the benefits of including video of the user's nonverbal interaction cues. There appears to be moderate agreement among evaluators that PIP is perceived as beneficial to identifying usability problems. As long as the cost for capturing user video remains inexpensive, most usability labs will include it unless there is data showing that it leads evaluators to find problems that are not useful. Future research in the Digital HCIL will attempt to quantify the benefits of including user video in usability evaluation recording sessions.

CONCLUSION

Our overarching goal was to develop a laboratory with equipment that was easy to use. Cadets using the laboratory equipment have developed expertise with the software and hardware in approximately 4 hours of training. Because the tools are so easy to use, cadets continue to find new ways of incorporating testing into various research projects. Starting in the Fall 2005 semester, we will use the laboratory in a new HCI course at the USAFA. During this course, cadets will spend about half of the semester in the laboratory, learning HCI methods and tools and conducting their own usability analysis on design projects. The HCI laboratory facilities are also being used by other organizations at the USAFA for evaluating local applications developed for cadets and faculty (e.g., web sites, management information systems, registration systems, etc.). In the future, our plan is to conduct research on improving usability evaluation methods and automating some of the techniques for usability problem identification.

Continued research in the Digital HCIL will also examine theoretical frameworks for identifying usability problems in a more objective manner. One study that will begin in the Fall 2005 semester is focused on integrating the User Action Framework developed at Virginia Tech with a Latent Semantic Analysis tool developed by Pearson Knowledge Technologies in Boulder, Colorado. These tools are being brought together as part of a small-business innovative research project sponsored by AFOSR. USAFA has been identified as a test site where these tools can be brought together and examined. Combining the User Action Framework (a problem identification taxonomy) with Latent Semantic Analysis (an automatic linguistic analysis technology) has the potential to bring more objectivity to the usability problem identification and diagnosis process and to enable usability engineers with less expertise to make faster and more accurate diagnoses. Research from the Digital HCIL at USAFA as a test site for these tools should provide an improved process for usability evaluation in the Air Force as well as industry.

REFERENCES

Argyle, M. (1972). Non-verbal communication in human social interaction. In R. A. Hinde (Ed.), *Non-verbal communication*. Cambridge: Cambridge University Press.

Argyle, M., & Dean, J. (1965). Eye-contact, distance and affilation. *Sociometry, 28,* 289-304.

Argyle, M., Lalljee, M., & Cook, M. (1968). The effects of visibility on interaction in a dyad. *Human Relations, 21,* 3-17.

Butler, K. A. (1996, Jan). Usability engineering turns 10. *Interactions, 3*, 58-75.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20(1),* 37-46.

Hix, D. & Hartson, H. R. (1993). *Developing User Interfaces: Ensuring usability through product & process*. New York: John Wiley & Sons, Inc.

Kendon, A. (1967). Some function of gaze-direction in social interaction. *Acta Psychologica, 26,* 22-63.

Landis, J. R. & Koch G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174.

Morae Overview Whitepaper (2004). Retrieved September 27, 2004, from http://www.techsmith.com/products/ morae/whitepaper.asp

Nielsen, J. (1993). *Usability engineering*. San Diego: Academic Press, Inc.

Patterson, M. L. (1983). *Nonverbal behavior: A functional perspective*. New York City, NY: Springer-Verlag.

Williges, R. C., & Hartson, H. R. (1986). Human-computer dialogue design and research issues. In R. W. Ehrich & R. C. Williges (Eds.), *Human-computer dialogue design* (pp. 367-376). Amsterdam: Elsevier.

Williges, R. C., Williges, B. H., & Elkerton, J. (1987). Software interface design. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 1416-1449). New York: Wiley.

ABOUT THE INSTITUTE

The Institute for Information Technology Applications (IITA) was formed in 1998 to provide a means to research and investigate new applications of information technology. The Institute encourages research in education and applications of the technology to Air Force problems that have a policy, management, or military importance. Research grants enhance professional development of researchers by providing opportunities to work on actual problems and to develop a professional network.

Sponsorship for the Institute is provided by the Assistant Secretary of the Air Force (Acquisition), the Air Force Office of Scientific Research, and the Dean of Faculty at the U.S. Air Force Academy. IITA coordinates a multidisciplinary approach to research that incorporates a wide variety of skills with cost-effective methods to achieve significant results. Proposals from the military and academic communities may be submitted at any time since awards are made on a rolling basis. Researchers have access to a highly flexible laboratory with broad bandwidth and diverse computing platforms.

To explore multifaceted topics, the Institute hosts single-theme conferences to encourage debate and discussion on issues facing the academic and military components of the nation. More narrowly focused workshops encourage policy discussion and potential solutions. IITA distributes conference proceedings and other publications nation-wide to those interested or affected by the subject matter.